DOCUMENT RESUME

ED 171 802 TM 009 509

AUTHOR Brandenburg, Dale C.
TITLE Some Statistical Properties of Item Specificity in Student Ratings.
PUB DATE Apr 79
NOTE 29p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (San Francisco, California, April, 1979)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Classification; Content Analysis; *Course Evaluation; Higher Education; *Item Analysis; Questionnaires; Rating Scales; *Student Evaluation of Teacher Performance; *Test Items; Test Reliability; Test Validity
IDENTIFIERS *Item Specificity

ABSTRACT

Prior research has indicated that items administered to college students for rating their instructors, can be empirically as well as logically classified on a continuum from very general to specific. Three of these hypothesized classifications of item specificity--global, general concept, and specific--were chosen to represent this continuum. Thirty-nine Likert-type items empirically identified as members of these categories, and as members of a content domain labeled influence and security, were then compared against six statistical properties: (1) skewness in distribution of class section means; (2) between-class variance of means; (3) within-class variance among student responses; (4) ceiling effect; (5) item reliability; and (6) interquartile range. Results indicated that most items met the criteria hypothesized, although some discrepancies for the most specific items were pronounced. The differentiation among specificity levels offered an essentially content-free classification scheme. Implications were drawn for questionnaire item writing, use of results, and the evaluation of overall item quality. (Author/CP)

Some Statistical Properties of Item Specificity in Student Ratings*

Dale C. Brandenburg

University of Illinois at Urbana-Champaign

Prior research has indicated that student rating items can be empirically as well as logically classified on a continuum from very general to specific. Three of these hypothesized classifications of item specificity (Global, General Concept and Specific) were chosen to represent this continuum. Items empirically identified as members of these categories were studied to discover further descriptive statistical behavior of item results and then tested against the selected items. Results indicated that most items met the criteria hypothesized although some discrepancies for the most specific items were pronounced. Implications were drawn for questionnaire item writing, use of results and evaluation of overall item quality.

*Paper presented at 1979 NCME Annual Meeting, San Francisco, California, April, 1979.

Some Statistical Properties of Item Specificity in Student Ratings

Dale C. Brandenburg

University of Illinois at Urbana-Champaign

The literature on students' ratings of college instructors and courses is abundant with correlational studies on how different variables influence ratings in addition to investigations of inter-relationships among the ratings themselves. Little attention, on the other hand, has been given to descriptive properties of items and subscales except for data included in technical reports of questionnaires (and these reports are not widely distributed). Furthermore, recent work by Brandenburg, Derry & Hengstler (1978) and Frey (1978), has indicated that one may have to examine more closely the specificity and general dimensions of items in order to draw conclusions from general correlational studies.

An important outcome of the Brandenburg, et al. (1978) study was that rating items could be empirically classified according to specificity. This classification by specificity for the most part was hypothesized by Smock and Crooks (1973) and by Rosenshine (1970). Specificity refers to the general to specific terminology used to word the items or alternatively, the amount of inference or judgment required on the part of students to respond to a given item. More generally stated items require higher inference or more judgment; more specifically or behaviorally stated items require less inference or less judgment. Labels for these items from most general to specific were given as Global, General Concept and Specific in the Brandenburg, et. al. (1978) study.

These levels of item specificity may also be related to how results from the different item types may be used. More general items provide information appropriately used in administrative (tenure, merit, etc.) decisions; specific items provide

information more appropriately used in giving feedback related to instructional improvement. Because general item results are used comparatively, norms are useful if not mandatory in order to derive less ambiguous interpretations. On the other hand, specific item results are not usually used comparatively so norms are not necessary, or in some cases would even be misleading.

This concept of specificity has been made potentially more complex with the recent work of Brandenburg, et. al. (1978) and Frey (1978). These studies make the claim for two higher-order domains of items in which Frey argues that the item members of one domain may be potentially more influenced by biasing effects such as class size, required/elective nature of course, and course level than are the item members of the second domain (Frey, 1978). Additionally, items in the second domain appear to correlate more highly with the student achievement and faculty publication records. Frey calls the first domain "Support" and the second "Pedagogical Skill"; in Brandenburg, et. al. (1978) the names were "Security" and "Influence," respectively. While the names of these two domains differ between the two studies and the items are slightly different, the constructs remain essentially the same. More importantly, the implications advocated for other studies is that conclusions such as "it was found that the student ratings differed significantly by course level" without having examined the higher-order domain to which such item(s) belong is probably improper and likely to be invalid for all items.

In the Brandenburg, et. al. (1978) study, it was shown that item specificity categories could be identified within each of these higher-order domains. Research using similar well-defined specific vs. more general items is sparse. Pohlman (1975) used primarily General Concept items to predict results on a Global item, and the outcome was generally positive. Cushman and Tom (1976) developed a set of "specific" teaching behaviors and correlated them to student perceptions of their progress on certain course objectives. While their resulting questionnaire contained a number of good psychometric qualities, its overall utility was limited due to the specific nature

of the questions. Rosenshine and Stevens (Note 1) correlated results on subgroups of specific items to more general questions related to each subgroup. Specific items that had low correlations with more general items were considered irrelevant for future considerations in student ratings. All of these studies were correlational and none attempted to define further descriptive characteristics of the general or more specific items.

If one assumes that the concept of item specificity has some practical utility there should exist some descriptive statistical properties of item types which may permit further item differentiation. Conversely, the assumptions regarding item specificity imply that certain statistical properties of items must exist for them to be classified in a given manner. For example, general items used for comparative purposes must have sufficient between-class-section variance in order to differentiate among instructors or classes. Specific items, on the other hand, might be hypothesized to show relatively small within-class variance, implying that students are responding to the same observed behavior, activity or trait. The objective of this investigation was to hypothesize some statistical properties related to item specificity and test their adequacy on some recently collected student rating data. A secondary purpose was to illustrate some descriptive characteristics of "good" items within a given classification. Such an examination should assist future student rating questionnaire developers in improving overall item construction and utilization procedures.

Definition of Terminology

Items selected for testing statistical properties were chosen from the Instructor and Course Evaluation System (ICES, Brandenburg, Note 2) Item Catalog. ICES is a flexible, computer-based CAFETERIA-type mechanism for collecting student ratings of instructors developed and used at the University of Illinois at Urbana-Champaign. The general parameters of this system permit instructors to select up to 23 items from the catalog which are computer-printed on a scannable answer sheet.

The ICES Item Cat[illegible]ontains about 450 items which are classified by conte[illegible] and item specificity. [illegible] the Brandenburg, et al. (1978) study it was found that items could be grouped into two large domains or halo dimensions labeled Influence and Security. In terms of item content, Influence consists primarily of items related to Student-Perceived Outcomes, Instructor Communication Skills, and Instructor/Course Stimulation or Motivation. Security item content consists primarily of items related to Course Management/Structure and Instructor Warmth/Concern. Course Difficulty or Workload is negatively related to Security and positively related to Influence

The interaction of the Influence and Security domains with item specificity is depicted in Figure 1 The large ellipse in the figure represents the total universe

- - - - - - - - - - - - - - - -
Insert Figure 1 About Here
- - - - - - - - - - - - - - - -

of student rating items. The two smaller ellipses within it represent the item domains of Influence Security which are further composed of Specific, General Concept and Global These labels for specificity represent the most specific to the most general respectively. Unique items are those which do not relate in any systematic manner to either content categories or general domains. Examples follow:

Global

#164--Do you feel course objectives were accomplished?

Yes, great extent	No, not at all

General Concept

#3 --The course was:

Organized	Disorganized

Specific

#265--The instructor made use of alternative explanations when needed.

Almost always	Almost never

Unique

#228--I went to sleep in class:

Very often ver

Global items are characterized as being equal or nearly equal members of both Influence and Security domains and their number is probably finite and small. General Concept items are seen as indicator items--results of which may illustrate general strengths or weaknesses of identified constructs such as speaking ability, course organization or sensitivity to students. Commonly these items would load highly on one factor in a factor analytic study defining a trait or construct in student ratings, whereas global items would typically load on two or more three factors. As opposed to Global and General Concept items, the number of Specific items is potentially infinite. Specific items, however should be related to a General Concept item within the same construct or factor; otherwise their utility is limited. This serves, then, to differentiate Specific from Unique items in that Unique items would not be related to any identifiable General Concept items (or factors).

Procedure

Data for the investigation was obtained from summary results for each class section utilizing ICES questionnaires during the 1977-78 academic year. Overall usage was about 5,000 class sections, but only 2,200 actually used the item catalog (instructors could also opt for pre-designed questionnaires).

All items used in this investigation contained five response positions with only end points labeled. Two item examples are:

#101--The grading procedures for the course were:
Very fair Very unfair

#241--Was the instructor a good speaker?
Yes, very good No, rather poor

Items chosen for this investigation were previously empirically identified in the Brandenburg et al., (1978), study as belonging to certain specificity cate-

gories and higher-order domains (Influence and Security). A list of the selected items within the classified dimensions is given in Table 1, and it includes 8 Global, 14 General Concept and 17 Specific items.

- - - - - - - - - - - - - - -
Insert Table 1 About Here
- - - - - - - - - - - - - - -

A second step of the procedure was to generate hypotheses concerning statistical properties which may be differentiable according to the previously delineated item types. Two sources of information were used to generate hypotheses--the theoretical distinction among item types related to item construction and purpose, and the experience of the author. Many of the hypotheses relate to an evaluation of item quality. For example, Global and General Concept items should have sufficient reliability to discriminate among instructors, but item reliability in this sense is irrelevant to Specific items.

The following loosely structured hypotheses were generated:

1. Skewness in distribution of class section means: Means for Global and General Concept items should be negatively skewed--Globals more so than General Concept. Specific item results are likely to be flatter, but also negatively skewed.

   Rationale: Negative skewness is a rather apparent characteristic of rating data for both within-class and between-class response distributions. With very few exceptions most questionnaires presently used contain almost exclusively Global and General Concept items, thus negative skewness is very commonly observed. While Specific items may also be negatively skewed, general halo effects should be less pronounced.

   Measures: $\bar{X}_B - Mdn_B$ (from overall item mean distributions)

   $$\left|C_{90} - Mdn\right| - \left|Mdn - C_{10}\right|$$

2. Between-class variance of means: Variance should be larger for Global than for General Concept. Specific items may be expected to have slightly smaller variance than general Concept items, although some Specific item variance may be exceptionally large.

   Rationale: Differentiation among classes or instructors is a primary intent of Global and General Concept items. The general wording of these items should permit students to respond to whether or not a given course or instructor has more or less of each trait rated. Comparisons among courses or instructors is an inevitable outcome accounting for substantial variance. Specific items, on the other hand, are theoretically only applicable in those classes where they were selected. This, itself, should account for a decrease in between-class variance. If Specific item variance is exceptionally large, it can probably be accounted for by a few extreme classes or instructors.

   Measure: s_B^2 (between-class variance of means)

3. Within-class variance among student responses: Specific item variance might be smaller than either Global or General Concept variance.

   Rationale: Specific items are meant to be behaviorally related to specific occurances, course materials, and instructor quirks observed or experienced in the classroom. If all students see the same thing and react the same way, within class variance should be small. For Global and General Concept items, the expectation for common student perception is unlikely.

   Measure: $\bar{s}_w^2$ (average within-class variance of responses for an item)

4. Ceiling effect: The negative skewness of Global and General Concept items produces little discrimination at the top of the distribution. Specific items should produce a lesser degree of this ceiling effect.

Rationale: To a certain extent the expectation for lower ceilings on Specific items is built on the premise that instructors will choose items corresponding to their weaknesses as well as to their strengths. To the extent that this is true, the hypothesis should be confirmed.

Measures: $(\text{Maximum} - X)_B$ Maximum weight minus mean for distribution as a whole.

$(\text{Maximum} - C_{90})_B$ Maximum weight minus C_{90} for distribution as a whole.

5. Item Reliability: Reliability should be largest for Globals (mid 80's), followed by General Concept (70's to mid 80's) and Specific (70's or lower).

Rationale: Item reliability as measured here is essentially an index of discriminating power, and it contains the ratio of within-class variance to between-class variance. Since between-class variance for Global and General Concept items is expected to be comparatively larger than that for Specific items, item reliability should be greater. This larger value of s_B^2 more than makes up for the smaller s_w^2 for Specific items.

Measure: Horst Reliability Formula

$$\text{rel} = 1 - \frac{\sum_1^N \frac{s_w^2}{n_w - 1}}{\frac{N}{s_B^2}}$$

where s_w^2 and s_B^2 defined as before

n_w = number of students within a class section

and N = number of class sections

6. Interquartile Range: Due to larger discriminating power of Global and General Concept items, the interquartile ranges of these items should be close to that of a normal distribution. The interquartile ranges of Specific items should be highly variable.

Rationale: This hypothesis assumes that Global and General Concept item mean distributions are approximately truncated normal curves. It is difficult to hypothesize a theoretical family of curves that Specific items would generally follow. In fact, they may not follow any family.

Measure: $Q_1 = C_{75} - C_{25}$

Results

Results of all descriptive indices or measures are displayed in Table 2 for each item. General Concept and Specific item results are presented separately for Influence and Security domains. In addition to the measures described above, it was

- - - - - - - - - - - - - - - -
Insert Table 2 About Here
- - - - - - - - - - - - - - - -

necessary for summary purposes to add further indices. The rows labeled s_B, $\bar{s}_W$, and $s_B^2/\bar{s}_W^2$ are self-explanatory. Measure ND/I refers to the ratio of the number of standard deviations in the interquartile range for a normal distribution (1.349) to the number of standard deviations (s_B) in the interquartile range of the mean distribution for a given item. Numbers greater than 1.0 for ND/I would thus indicate more spread in a normal distribution for the middle 50% of class means than for the item under study. The last row labeled "N" refers to the number of class sections in which each item was selected. It should also be noted that in the results for Influence-Specific items 19 and 144 and Security-Specific items 24, 52, and 122, the mid-point was the most positive response. Thus measures $\bar{s}_W^2$, s_B^2, and Q_1 are probably deflated. "Max" was chosen to be 3.0 so $(\text{Max} - \bar{X})_B$ has a different meaning from other items and $(\text{Max} - C_{90})$ has no meaning.

Table 3 was constructed to assist in summarizing the data in Table 2. Table 3 contains the medians of selected indices from Table 2 data. Results are presented according to hypothesis.

- - - - - - - - - - - - - - - -
Insert Table 3 About Here
- - - - - - - - - - - - - - - -

1. Skewness in Distributions of Class Section Means

The first measure of skewness ($\bar{X}_B - Mdn_B$) yielded only minor differences among item types. Most measures were around -.10 or less. Results for the second measure $|C_{90} - C_{50}| - |C_{50} - C_{10}|$ did indicate that Global items were slightly more negatively skewed than General Concept items (see Table 3). Specific items, on the other hand, fluctuated from slight positive skewness to very negative. This result may in part be due to the smaller number of class sections using Specific items.

2. Between Class Variance of Means (s_B^2)

Except for item #170 there was only slight variability of s_B^2 for Global items. The average s_B^2 for Influence-General Concept items was higher than for any other item grouping (see Table 3). Specific item s_B^2 varied considerably among items (range .121 to .828 for non-mid-point items), and this unpredictability was expected. If the items selected are representative of "good" Global and General Concept items an s_B of .50 appears to indicate a criterion of quality.

3. Within-Class Variance Among Student Responses ($\bar{s}_w^2$)

The average within-class variability ($\bar{s}_w^2$) was found to be lower for Global items than any other item group. This was contrary to expectations. Influence-General Concept items yielded the highest average s_w^2 and this was consistent with expectations. It can also be noted from Table 2 that with the exception of item #200, all Global and General Concept items s_w^2 were between .67 and .87--a high degree of consistency. The finding that $\bar{s}_w^2$ for Specific items was little different from $\bar{s}_w^2$ for other items suggests a rethinking for expectations of Specific items and their associated utility. At least the mid-point, best response Specific items yielded the lowest values of $\bar{s}_w^2$ as might be expected.

4. Ceiling Effect

The influence of greater skewness of Global items also probably accounts for the greater ceiling effect observed for the measures (Max - $\bar{X}$) and (Max - C_{90}). For Global items the room at the top (Max - $\bar{X}$) of the class means distributions are

generally less than two standard deviations (s_B) while for all other item groups the distance from the maximum to the mean is at least two s_B. It is also worthwhile to note that Influence-General Concept item means were substantially lower (about 1/2 s_B) than means from all other item groups.

5. Item Reliabilities

In general there were no differences between Global and General Concept items for reliability. Except for item #170 the range was .69 to .91, and a median of .80 was observed for each subgroup. A value of .80 may be a useful criterion for items of these types because discrimination power is needed. The reliability for Specific items on the other hand, fluctuated a great deal and conformed to the hypothesized low .70's. This may be due in part to the smaller sample sizes on which Specific item data was based.

6. Interquartile Range

The results for the interquartile range (Q_1) conform to those given for ceiling effect. That is, Influence-General Concept items had a substantially larger Q_1 than did other item groups (see Table 3). In order to get a better idea of how Q_1 compares to that expected in a normal curve, the measure ND/I was determined. The values as given in Table 2 show that Global and General Concept items provide about as much discrimination in the center of the means distributions as a normal curve. The range of ND/I with one exception (item #200) was .92 to 1.1. One-half of the General Concept items had values less than 1.0 indicating slightly more spread than that for a normal distribution. As may be anticipated, the spread of ND/I values for Specific items is much larger than that for other item groups, and the number of values below 1.0 (7) was exceeded by the number above 1.0 (10).

Discussion

The differentiation among specificity levels of student rating items offers an essentially content-free classification scheme. It also has some potential long run

benefits in the evaluation and improvement of college instruction. In order for these benefits to accrue, item specificity implications have to be thoroughly investigated. This study represents an attempt to articulate some descriptive statistical indices related to item specificity implications.

To a large extent, the path for investigating these implications directly interacts with the investigation of item quality. To illustrate, for Specific items to have maximum utility in instructional feedback situations, one might assume that within-class variance be small to assure that faculty development efforts are pointed in the proper direction. Along the same line, Specific items should be worded so that growth may be observed over time--high ceilings on item means would prohibit this observation. Thus, if Specific items do not yield certain statistical properties related to their theoretical purpose, then more work has to be done to improve the items or they should be eliminated from use in the classroom.

Interpretation of the above results for Specific items in this context leads one to question the overall quality of such items selected for this study or to rethink the hypotheses regarding the statistical behavior of these items. More specifically, $\bar{s}_w^2$ was higher than anticipated (about equal to $\bar{s}_w^2$ for other item groups) and the ceiling effect was just as prevalent if not more so for Security-Specific items. The ceiling problem may be due in part to instructors selecting items they know students will give high ratings. But since results of such items are not routinely sent to department heads, this behavior does not appear to be highly rational. On the other hand, the statistical properties selected for use in this study may not have been completely fair to judge Specific item quality. Perhaps more attention to within-class distributions would have yielded more positive results.

Taking another approach, it may be useful to hypothesize alternate explanations for the observed discrepancies. This explanation involves six potential reasons[1]

[1]The author is indebted to H. Richard Smock for his thoughts on this topic.

which may account for the observed data. (1) Global items are easier to answer; they contain a halo effect that Specific items do not possess. This may account for the lower $\bar{s}_w^2$ obtained for Global items. (2) Responses to all items include learner differences, cognitive differences and differences in approaches to learning. These differences may serve to increase variability on Specific items, not decrease it. (3) The semantics of the responses may be more of a factor in responding to Specific items than to other item types. (4) Recency effects, those occurrences or interactions observed or felt in the classroom situation impinging upon the student at the time he/she is responding to a question, may influence results on Specific items more so than on others. (5) There exist affective or emotional differences among students (regardless of the homogeneity of general intellect) which Specific items may be more likely to elicit that would also serve to increase variability. (6) Specific items might also trigger associations with authority and past comparitors, such as former teachers, parents or other acquaintances, more so than other items. While these six are conjectures not generally researchable, they at least provide a framework which future investigations may take into account.

Parallel considerations for item quality apply to Global and General Concept items. Prior arguments have been made in this case for between-class variance and item reliability. If the items do not satisfy these criteria, they also must be reworded or eliminated because discrimination among instructors or classes is a primary function of these items. Such criteria include an $s_B \geq .5$ (5 point scale), Rel $\geq .80$, $(\text{Max} - \bar{X}) \geq 2s_B$ and ND/I ≤ 1.1. These criteria would appear to permit adequate differentiation among class means. It is quite probable that these criteria are not standards, but they do permit a starting point for judging item quality. Most Global and General Concept items included in this study met and surpassed these criteria.

It is also worthwhile to note that Influence-General Concept items appeared to behave quite differently from Security-General Concept items. Notable differences

included an s_B 40% larger for the Influence subset and a (Max - $\overline{X}$) range that was about 1/2 s_B larger. Part of these differences may be due to the item sample, but most may be accounted for by the general content domains represented in the items. Maybe the larger s_B and (Max - $\overline{X}$) indices observed are a product of the less susceptibility to biasing effects of Influence items. This is yet another topic for further study.

The experience gained in this study is transferable to the investigation of other items in the ICES catalog as well as to the examination of other questionnaires. If item purpose (use) is specified together with a judgment of general to specific wording used, then some examination of item quality can proceed. The methods presented here provide a start for such efforts which in the long run should result in better quality student rating data and results more likely to withstand the pressures for appropriate use.

Reference Notes

1. Rosenshine, Barak and Robert Stevens. Specific college teaching behaviors. Mimeographed. University of Illinois College of Education, undated.

2. Brandenburg, Dale C. ICES: Its Rationale and Description. ICES Newsletter No. 2, Office of Instructional Resources, University of Illinois, August, 1977.

References

Brandenburg, Dale C., Sharon Derry and Dennis D. Hengstler. Validation of an Item Classification Scheme for a Student Rating Item Catalog. Paper presented at the NCME Annual Meeting, March, 1978. Toronto, Canada (ERIC TM007169).

Cushman, Harold R. and Frederick K.T. Tom. The Cornell Diagnostic Observation and Reporting System for Student Description of College Teaching. NACTA Journal, March, 1976, 10-16.

Frey, Peter W. A two-dimensional analysis of student ratings of instruction. Research in Higher Education, 1978, 9, 69-91.

Pohlmann, John T. A description of teaching effectiveness as measured by student ratings. Journal of Educational Measurement, 1975, 12, 49-54.

Rosenshine, Barak. Enthusiastic teaching: A research review. School Review, 1970, 78, 499-514.

Smock, H. Richard and Terence J. Crook. A plan for the comprehensive evaluation of co-lege teachers. Journal of Higher Education, 1973, 44, 577-586.

Figure 1

Hierarchical Schematic Classification

of Student Rating Items

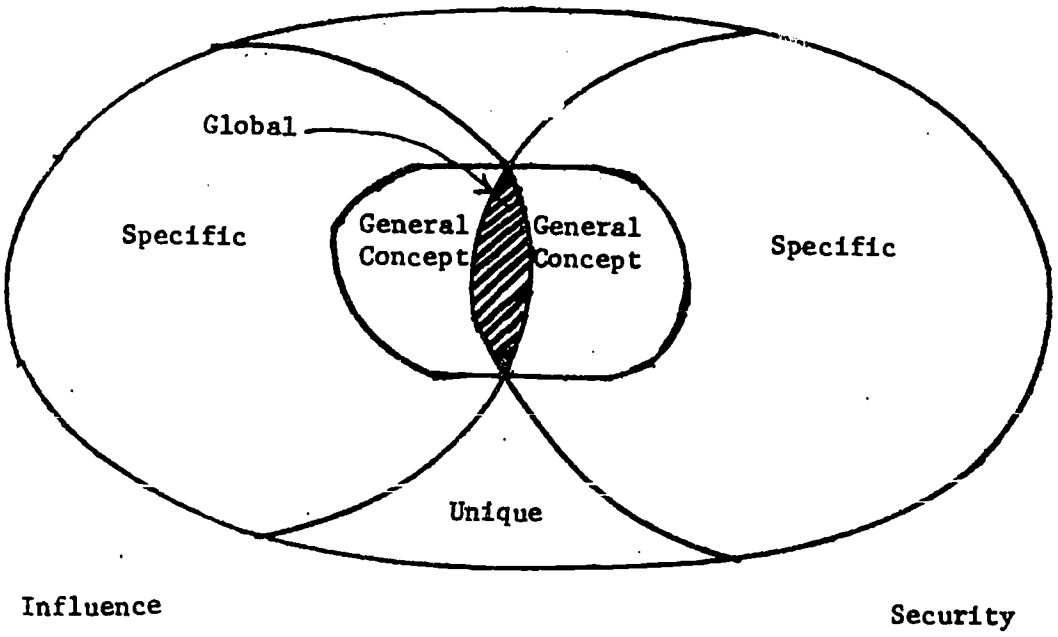

Table 1

Selected Items and Their Hierarchical Classification

GLOBAL

2--The instructor stated clearly what was expected of students.

Almost always	Almost never

5--Was the progression of the course logical and coherent from beginning to end?

Yes, always	No, seldom

13--Was class time spent on unimportant and irrelevant material?

Yes, often	No, never

164--Do you feel course objectives were accomplished?

Yes, to a great extent	No, not at all

169--Did this course improve your understanding of concepts and principles in this field?

Yes, significantly	No, not much

170--Can you now identify main points and central issues in this field?

Yes, clearly	Not very well

195--Did your interest in this course increase or decrease as the semester progressed?

Greatly increased	Greatly decreased

240--The instructor was a dynamic teacher.

Yes, very dynamic	No, very dull

INFLUENCE-GENERAL CONCEPT

46--How would you rate instructional materials used in this course?

Excellent	Poor

160--How much do you feel you have accomplished in this course?

A great deal	Very little

162--How much have you learned in this course?

A great deal	Very little

200--Were you stimulated to discuss related topics with friends outside of class?

Yes, often	No, never

204--I developed a more positive self-concept because of this course.

To a great extent	Not at all

220--Compared to other courses, how much effort did you put into this course?

Much more	Much less

255--How interesting were the instructor's presentations?

Very interesting	Rather boring

325--The instructor motivated me to do my best work.

Almost always	Almost never

SECURITY-GENERAL CONCEPT

3--The course was:

Organized	Disorganized

4--Was there agreement between announced course objectives and what was taught?

Strong agreement	No agreement

Table 1

(continued)

101--The grading procedures for the course were:
Very ... Very unfair

105--Did the instructor have a realistic definite excellent performance?
Yes, ... No, very
realistic ... unrealistic

286--The instructor's presentation of abstract ideas, concepts, and theories was:
Very clear ... Very unclear

362--The instructor seemed to sense when students did not understand.
Strongly agree ... Strongly disagree

INFLUENCE-SPECIFIC

19--The course content was:
Too advanced ... Too elementary

50--Were readings well selected?
Yes, all very good ... No, all very poof

63--Describe your written assignments.
Interesting stimulating ... Dull, uninspiring

116--Did the exams challenge you to do original thinking?
Yes, very challenging ... No, not challenging

144--Describe the pace of the course.
Too fast ... Too slow

328--Did the instructor raise challenging questions in class?
Yes, often ... No, seldom

335--Did the instructor encourage you to develop your ideas and approaches to problems?
Definitely yes ... Definitely no

382--Was a good balance of student participation and instructor contribution achieved?
Always ... Never

SECURITY-SPECIFIC

24--Should more/less time be provided to review and synthesize course material?
Much more time ... Much less time

52--Did the readings require a reasonable amount of time and effort?
No, too demanding ... No, too simple

114--The exams reflected important points in the reading assignments.
Strongly agree ... Strongly disagree

122--How difficult were the examinations?
Too difficult ... Too easy

265--The instructor made use of alternative explanations when needed.
Almost always ... Almost never

340--Did the instructor suggest specific ways students could improve?
Yes, frequently ... No, almost never

354--The instructor listened attentively to what class members had to say.
Always ... Seldom

378--Was the instructor cynical and sarcastic?
Very cynical ... Not at all cynical

381--In terms of direction and structure of the course, the instructor was:
Flexible ... Rigid

Table 2

Descriptive Statistics for Selected ICES Items

Item No.	GLOBAL								INFLUENCE-GENERAL CONCEPT							
	2	5	13	164	169	170	195	240	46	160	162	200	204	220	255	325
$\bar{X}_B - Mdn_B$	-.09	-.05	-.07	-.10	-.07	-.08	+.01	-.10	-.01	-.07	-.05	-.10	-.09	+.02	-.10	-.04
$(C_{90}-C_{50})_B - (C_{50}-C_{10})_B$	-.25	-.12	-.11	-.30	-.21	-.25	+.04	-.27	-.07	-.19	-.14	-.65	-.49	+.02	-.32	-.16
s_B	.51	.52	.47	.48	.55	.38	.60	.66	.52	.56	.45	.76	.61	.58	.69	.61
s_B^2	.263	.267	.223	.226	.302	.143	.355	.431	.266	.313	.199	.578	.373	.335	.474	.372
$\bar{s}_W$	.73	.74	.71	.67	.81	.82	.70	.74	.80	.85	.81	1.03	.82	.83	.84	.87
$\bar{s}_W^2$	.533	.548	.504	.449	.656	.672	.490	.548	.640	.723	.656	1.06	.672	.689	.706	.757
$s_B^2/\bar{s}_W^2$	.49	.49	.44	.50	.46	.21	.72	.79	.42	.43	.30	.55	.56	.45	.67	.49
$(Max-\bar{X})_B$	.80	.91	1.13	.84	.91	.97	1.64	1.04	1.19	1.07	.91	1.52	1.32	1.47	1.24	1.27
$(Max-C_{90})_B$	.21	.28	.54	.23	.22	.58	.89	.24	.58	.39	.38	.73	.70	.70	.45	.46
$C_{75}-C_{25\ B}$	.68	.64	.62	.60	.74	.56	.81	.90	.63	.80	.58	.82	.86	.80	.92	.83
MD/I	1.0	1.1	1.0	1.1	1.0	.92	1.0	.99	1.1	.94	1.0	1.2	.96	.98	1.0	.99
Rel	.799	.809	.678	.766	.799	.483	.744	.910	.749	.786	.734	.867	.759	.812	.887	.834
N	786	455	127	175	691	122	183	336	345	435	507	70	54	682	204	575

Table 2 (Continued)

Descriptive Statistics for Selected ICES Items

Item No.	SECURITY-GENERAL CONCEPT						INFLUENCE-SPECIFIC							
	3	4	101	105	286	362	19*	50	63	116	144*	328	335	382
$X_B - Mdn_B$	-.04	-.06	-.04	-.05	-.05	-.08	+.03	-.07	+.04	-.01	-.06	-.10	-.04	-.06
$C_{90}-C_{50\ B}$ - $C_{50}-C_{10\ B}$	-.11	-.13	-.16	-.14	-.18	-.24	+.14	-.17	+.19	+.05	-.32	-.28	-.08	-.18
s_B	.51	.46	.53	.50	.50	.43	.29	.51	.47	.35	.39	.48	.49	.53
$\bar{s}_B^2$	.265	.211	.276	.251	.255	.187	.086	.257	.223	.121	.151	.232	.244	.276
$\bar{s}_W$	.75	.70	.82	.89	.78	.83	.57	.74	.92	.82	.60	.82	.80	.83
$\bar{s}_W^2$	.563	.490	.672	.792	.608	.689	.325	.548	.846	.672	.360	.672	.640	.689
$s_B^2/\bar{s}_W^2$	.47	.43	.41	.32	.42	.27	.26	.47	.26	.18	.42	.35	.38	.40
$(Max-\bar{X})_B$	.87	.79	1.03	1.19	1.16	1.05	-.14	1.07	1.46	.81	-.21	.95	.92	1.20
$(Max-C_{90})_B$	.24	.24	.39	.55	.50	.55	x	.45	.81	.34	x	.40	.32	.55
$C_{75}-C_{25\ B}$	.72	.59	.72	.68	.62	.58	.35	.65	.50	.53	.46	.67	.61	.73
MD/I	.96	1.1	.99	.99	1.1	1.0	1.1	1.1	1.3	.89	1.1	.97	1.1	1.0
$\bar{r}_{ei}$	.812	.801	.787	.738	.828	.690	.547	.720	.722	.695	.736	.776	.746	.816
N	525	473	1031	210	194	237	364	111	79	71	253	216	125	352

*These items were scored so that the most positive response was '3' rather than 5 for other items.

Table 2 (continued)

Descriptive Statistics for Selected ICES Items

	SECURITY-SPECIFIC								
Item No.	24*	52*	114	122*	265	340	354	378	381
$\bar{X}_B - Mdn_B$	+.02	-.05	-.34	-.02	-.05	-.10	-.11	-.22	-.02
$(C_{90} C_{50})_B -$ $(C_{50} - C_{10})_B$	-.07	-.09	+.10	.00	-.15	-.22	-.30	-1.00	-.04
s_B	.35	.29	.67	.36	.41	.58	.42	.91	.36
s_B^2	.124	.084	.450	.129	.168	.332	.173	.828	.127
$\bar{s}_W$	.71	.65	.90	.64	.77	.78	.69	.80	.79
$\bar{s}_W^2$	.504	.423	.810	.410	.593	.608	.476	.640	.624
$s_B^2/\bar{s}_W^2$	.24	.20	.56	.31	.28	.55	.36	1.3	.20
$(Max-\bar{X})_B$	-.60	-.44	1.35	-.40	.95	.88	.47	1.04	.89
$(Max-C_{90})_B$	x	x	.60	x	.46	.17	.06	.14	.36
$C_{75}-C_{25\ B}$	.46	.45	1.12	.46	.55	.85	.49	1.45	.54
ND/I	1.0	.87	.81	1.1	1.0	.92	1.2	.85	.90
Rel	.678	.643	.845	.814	.685	.796	.723	.920	.560
N	163	92	49	150	97	98	247	23	47

*Items were scored so that the most positive response was '3' rather than 5 for other items.

Table 3

Medians of Selected Descriptive Indices

Item Classification	$\bar{X}$-Mdn	$(C_{90}-C_{50})-(C_{50}-C_{10})$	Interquartile Range*	Max-$\bar{X}$*	Max-C_{90}*	$\bar{s}^2_W$*	s^2_B	Rel	$s^2_B/\bar{s}^2_W$
Global	-.07	-.23	.66	.94	.28	.54	.27	.80	.49
Influence General Concept	-.05	-.17	.81	1.26	.52	.70	.35	.80	.49
Security General Concept	-.05	-.15	.65	1.04	.44	.65	.25	.80	.41
Influence Specific	-.05	-.12	.63	1.01	.42	.68	.23	.73	.36
Security Specific	-.05	-.09	.55	.95	.26	.62	.17	.72	.31

*Items with a most positive response of 3.0 were omitted from these measures.